ABSTRACT

METHOD OF, SYSTEM FOR, AND COMPUTER PROGRAM PRODUCT FOR CREATING AND CONVERTING TO UNICODE DATA FROM SINGLE BYTE CHARACTER SETS, DOUBLE BYTE CHARACTER SETS, OR MIXED CHARACTER SETS COMPRISING BOTH SINGLE BYTE AND DOUBLE BYTE CHARACTER SETS

Methods for specifying the types of constants whose character values are to be converted to Unicode; for specifying which code page or pages are used for specifying the character encodings used in the source program for writing the character strings to be converted to Unicode; and that can be used to perform conversions from SBCS, mixed SBCS/DBCS, and pure DBCS character strings to Unicode. A syntax suitable for specifying character data conversion from SBCS, mixed SBCS/DBCS, and pure DBCS representation to Unicode utilizes an extension to the conventional constant subtype notation. In converting the nominal value data to Unicode, currently relevant SBCS and DBCS code pages are used, as specified by three levels or scopes derived from either global options, from local AOPTIONS statement specifications, or from constant-specific modifiers. Global code page specifications apply to the entire source program. These global specifications allow a programmer to declare the source-program code page or code pages just once. These specifications then apply to all constants containing a request for conversion to Unicode. Local code page specifications apply to all subsequent source-program statements. These local specifications allow the programmer to create groups of statements containing Unicode conversion requests, all of which use the same code page or code pages for their source-character encodings. Code page specifications that apply to individual constants allow a detailed level of control over the source data encodings to be used for Unicode conversion. The conversion of source data to Unicode may be implemented inherently to the translator (assembler, compiler, or interpreter) wherein it recognizes and parses the complete syntax of the statement in which the constant or constants is specified, and performs the requested conversion. Alternatively, an external function may be

STL9-2000-0055   46

1   invoked by a variety of source language syntaxes which parses as little or as much of the source

2   statement as its implementation provides, and returns the converted value for inclusion in the

3   generated machine language of the object program. Alternatively, the conversion may be

4   provided by the translator's macro instruction definition facility.